

PRESSDOK – Micro-ISIS kumulációban

Felhasználói tapasztalatok a szegedi Egyetemi Könyvtárban

Kokas Károly

Előzmények

Hallgatóink és oktatóink általános társadalomtudományi érdeklődése hívta létre úgy 1975 táján a magyar nyelvű politikai, gazdasági, kulturális sajtó legjavából készített házi katalógusunkat. A válogatás fő forrása az MNB Időszaki Kiadványok Repertóriumra volt, amelyet egy rövid időszakban kiegészítettünk a Szabó Ervin Könyvtár hasonló tematikájú cédula-szolgáltatásával. A nyolcvanas évek közepére a tárgyszavak szerint rendezett cédulamennyiség (*társadalomtudományi cikk-katalógus*) mintegy 20-25 ezer magyar folyóirat- és újságcikkről adott információt, az idők folyamán maximálisan a nálunk jelentkező olvasói igényekhez igazodva.

Amikor a szegedi Egyetemi Könyvtár megvásárolta az első Commodore 64 típusú gépeket, magától értetődően felvetődött, hogy a mikroszámítógépes adatbázisépítés lehetőségeit egy viszonylag kisebb, csekélyebb évi gyarapodású, de a legfontosabb keresési, nyomtatási problémákat mégis felvető anyag gépesítésén próbáljuk ki. A könyvtári nagy-katalógus gépesítése ekkor már jó tíz éve egészen más pályán mozgott, így ennek a gépkategóriának a bevonása ott fel sem merült.¹ A Commodore-ral az elérhető legnagyobb kapacitású lemezmeghajtót használtuk, az 1 Mbájtos SD 1001 típusút. A felhasznált szoftver (Superbase 64) nem tette lehetővé a magyar ékezetes adatbevitelt, ennek ellenére a szolgáltatás népszerűsége jelentősen megnőtt, nem kis mértékben annak újszerű volta miatt (kinyomtatott listán kapta az olvasó az irodalomjegyzéket, források szerint rendezve). Mintegy 4000 rekordot tartalmazott adatbázisunk, amikor lehetőség adódott, hogy IBM PC kompatibilis gépre térjünk át. Az anyagot – a JATE Kalmár László Kibernetikai Laboratórium munkatársai segítségével – dBASE formátumra konvertáltuk, majd az első CDS/Micro-ISIS verzió elérésével rögtön az ISIS-be. Ebben a két utóbbi formátumban kísérletet tettünk az adatbázis floppyra való terjesztésére, pár hónap alatt nem túl nagy reklámmal 5 példányt tudtunk értékesíteni. Eközben egyre nagyobb nehézséget jelentett, hogy kapacitáshiány miatt csak állandó késéssel tudtuk az anyagot rögzíteni, jórészt amiatt is, hogy a megnövekedett érdeklődés nyomán újabb és újabb lapokat vontunk be a feldolgozásba. A gordiuszi csomót

az Országgyűlési Könyvtár PRESSDOK szolgáltatásának megismerésével és előfizetésével vágtuk át.

A társadalomtudományi cikk-katalógus és a PRESSDOK

Nem volt nehéz rájönnünk, hogy a konkurencia nem csupán kongeniális, hanem gyorsabb is a mi szolgáltatásunknál, és jóval nagyobb merítésű.² Természetesen előzetes eredményeink birtokában bennünket nem az eredetileg kínált floppy-alapú, önálló keresőszoftverrel felszerelt változat érdekelt, hanem a már meglévő adatinkkal kumulálható, az ISIS számára "fogyasztható" ISO-formátum. Speciális igényeink is voltak, hiszen alkalmazkodni szerettünk volna az Egyetemünkön elterjedt karakterkiosztáshoz, hogy az átégetett EPROM-mal rendelkező gépeken és nyomtatókon a megjelenítés és rendezés jól működjön. Az Országgyűlési Könyvtár kitűnő munkatársaival igen gyorsan szót értettünk, így 1989 eleje óta a PRESSDOK mágneslemezen érkezik könyvtárunkba. A kapott ISO-formátumot (amely már a kívánságunk szerinti karakterkonverzióval érkezik) így már egyszerű *import FST* segítségével a saját régebbi rekordszerkezetünkbe be tudjuk tölteni.

Kumulált PRESSDOK Micro-ISIS-en

A további adaptáció során olyan apróbb problémákat kellett megoldanunk, amelyek az ISIS-felhasználás általánosabb tanulságai irányába mutatnak.

- ▶ Eredeti adatbázisunk az alkalmazott szoftver miatt ékezet nélküli volt, viszont a természetes írásképnak megfelelően nagy- és kisbetűs.
- ▶ A kapott ISO-fájlok ékezhelyesen ugyan, de csupa nagybetűs változatban érkeznek hozzánk (feltehetőleg az eredeti szolgáltatás számára így készülnek).
- ▶ Mi eredetileg hagyományos hierarchizált tárgyszórendszerrel dolgoztunk, a PRESSDOK deskriptor-lánccal írja le a dokumentumok tartalmát.

- ▶ Eddigi rendszerünkhöz képest viszont teljesen új volt a PRESSDOK önálló kódrendszere, amely egyfajta szakrendi jelzetként szolgál.

Az ISIS által nyújtott lehetőségeket figyelembe véve, célszerűnek mutatkozott az adatbázis-hoz rendelt két karaktertábla (ISISUC és ISISAC) oly módon történő átírása, hogy keresési fájl számára invertáláskor egységesen ékezet nélküli nagybetűs formára konvertálja a kereső-kifejezéseket. Ezzel ugyan elvesztettük a kereséskor alkalmazható "finomítások" egy részét, viszont nem kell minden keresést kétszer fogalmazni, ti. egyszer az ékezetes, egyszer pedig az anélküli kifejezésekre. Ezzel megoldódott a vegyes írásmód és a csupa nagybetűs íráskép közös konverziója. Természetesen sort/print opciókban a karakterhelyes sorrend érvényes továbbra is.

A PRESSDOK deskriptor-sorát a mi tárgyszó-mezőnkbe konvertáltuk, ami azzal járt, hogy az egész mezőt szavanként invertáltatjuk, így a korábbi tárgyszavak minden külön írt eleme és a mostani leíró rész minden eleme egy-egy lehetséges kereső-kifejezést alkot. A PRESSDOK-ban azonban ez utóbbi mező alkalmanként kitöltetlen, mivel a cím maradéktalanul kifejezi a tartalmat, és a rendszert eredetileg szekvenciális keresésre specifikálták. Ezt a problémát csak úgy tudtuk volna korrekt módon áthidalni, ha utólag deskriptorozzuk a "hiányos" rekordokat. Egyelőre praktikus megfontolásból a cím-mező teljes invertálása mellett döntöttünk. Az immár két mező szavankénti feldolgoztatása kényelmes megoldásnak tűnhet, ám további kérdéseket vetett fel. A rendszer maximális lehetőségeit kihasználó tiltott szójegyzéket (STOPWORD-LIST) kellett létrehozni, amely a lehető legkisebbre korlátozza a névelő-, névutó-, módosítószó-dzsungel szaporodását a kereső-kifejezések szótárában (TERM DICTIONARY). A több kísérlet után létrehozott kb. 950 szavas szójegyzék számításaink szerint több tízezer felesleges, semmitmondó kifejezéstől szabadít meg, jelentékeny sebességnövekedés mellett.

A PRESSDOK említett kódrendszere első pillanatban kissé egyedi, de nem haszontalan munkaeszköznek tűnt. Zavaróak voltak egyrészt hiányosságai (néhány terület kódja kimaradt), másrészt túlságosan nagy témákat fedett le egy-egy kód. Előnyére vált viszont, hogy két-három karakteres lévén, még teljes invertálás esetén sem "falt fel" túl sok helyet a winchestereken. Sajnos

az Országgyűlési Könyvtár a PRESSDOK-ot – a több éves kumulációban használókkal nem számolva – 1990 elejével megváltoztatta a kódrendszert, mégpedig úgy, hogy a korábbi azonosítók egy részét újra kiosztotta, csak most más témakörre. Így kereséskor ugyanaz a kód más tartalmakat gyűjt ki a '89-es és ismét másokat a '90-es anyagból. Külön gondot jelent, hogy a *magyar*, *Magyarország* fogalom általában nincs kóddal ellátva, így csak negatív módon szűrhető ki a csak hazai vonatkozású, de amúgy általános témák, pl. a választások közül a magyar választásokra vonatkozó irodalom. Ez még ANYFILE alkalmazásával is igen körülményes. A kódok okozta problémát egyelőre nem tudtuk áthidalni, így jelenleg a PRESSDOK tartalmi kódjaira a rendszerünk nincs indexelve. Ha több ilyen "reform" nem következik, akkor esetleg később az 1990 előtti anyag külön kezelésével orvosolható lesz ez a gond, jelenleg azonban erre még nincs mód, mivel felhasználóink, különösen a több éves kumuláció miatt kedvelik a szolgáltatást.

A természetes nyelvű mezők (cím és tárgyszó/deszriptorlánc) szavankénti invertálása természetesen a különböző jellemző tartalmak igen nagy szétszóródását eredményezte. Ezt a nehézséget egy folyton továbbfejlesztett és hatékonynak mondható ISIS-módszerrel próbáltuk megoldani: a leggyakrabban keresett témák lehetséges elérési pontjainak megfelelően létrehoztunk egy olyan kereső–szinoníma fájlt (ANYFILE), amely pl. a földrajzi területek szerinti keresést nagyban megkönnyíti. Így pl. az *any kelet-európa* kifejezés keresőkérdésben való alkalmazása minden kelet-európai államot főnévi és melléknévi alakban "meghív" (Lengyelország, lengyel, bolgár, bulgár, stb.) a csonkolt változatú alakban is.

Az ISIS egy különösen kellemetlen, főleg ennyi szabad szövegből vett kifejezés esetén zavaró tulajdonságán némi programozói segítséggel tudtunk túljutni. Ismeretes, hogy nagy keresési szótár esetén a túl korán csonkolt kereső kérdést az ISIS órákig tartó bogarászással "hálálja meg". Ennek legfontosabb oka, hogy a program nagyszámú közbülső találat esetén visszaír a winchesterre, hogy legyen miből a találati fájlt megkonstruálnia. Ez az eset néha nem látható előre, néha pedig nem elkerülhető (pl. a *szovjet*-típusú csonkolt keresés igen gyakori). Említett kollégáink olyan kis segédprogramot dolgoztak ki (természetesen ISIS-Pascalban), amivel az

adatbázis statisztikailag villámgyorsan megszondázható arra vonatkozóan, hogy érdemes-e szűkebb csonkolással belevágni a keresésbe.

Felhasználói tapasztalatunk alapján újabban nem invertáljuk a cikkek szerzőit sem, azon egyszerű oknál fogva, hogy igen ritkán keresik e szempont szerint az információt. Ha mégis, akkor valamilyen szempontból már leválogatott keresési unióban (SET) élünk az ISIS 2.0 változat óta meglévő szabad szöveg visszakereséssel (FREE TEXT SEARCHING), természetesen a megadott rekordokon belül és csak a szerzői (vagy az éppen kívánt egyéb) mező(k)ben.

Hardver- és szoftver-tapasztalatok

A saját korábbi rekordjainkkal kumulált PRESSDOK-adatbázisunk a cikk írásának időpontjában csaknem 50000 rekordot számlál, az elérési pontok száma (POSTING) mintegy 400000. Maga az alapfájl jelenleg 9.5 Mbájt, a tényleges invertált fájlokkal együtt ez kb. 20 Mbájt. Az invertáláskor keletkező ún. link fájlokat belső útvonal tereléssel (SYSPAR.PAR) eleve a gép másik winchesterén építetjük, külön al-könyvtárban, hiszen azok a valódi invertált fájl felépülése után törölhetők. Ez nem jelentéktelen mennyiség, a fenti rekordszám mellett kb. 21-22 Mbájt.

A PRESSDOK-ot kezdettől fogva IBM AT típusú (Intel 80286 12 mhz) gépen forgalmazzuk, a gép mindkét (formatáltan) 32 Mbájt kapacitású winchesterét igénybevéve. Ez a gyakorlatban azt jelenti, hogy a havi kb. 2500 rekord megérkezésekor – az invertálás idejére – csaknem minden más lekerül a gépről, majd a folyamat végeztével visszaköltözik. Ez persze már a mai korszerű backup programok mellett percek kérdése csupán. A két említett mezőre irányuló index-készítés még egy ilyen viszonylag gyorsabb gépen is rendkívül időigényes, ezért nem a teljes invertálást választjuk, hanem az ISIS által kínált három szakaszos megoldást ('I' menü: G,S,C opciók). E megoldás előnye, hogy egy-egy szakasz a reggeli (kevésbé forgalmas) órákban lefut, és az utolsó lépést kivéve (az új invertált fájl betöltése) a korábbi index végig használható. A három napra elnyújtott aktualizálási művelettől így módon megkíméljük az olvasókat és némileg a gépet is.

A PRESSDOK és az olvasók

A szolgáltatás nem túl nagy sajtó- és plakát-reklám után indult ebben a formában Szegeden. A gyakorlati módszer az online közvetítői praxis-hoz hasonló. A keresést végző kolléga akár kísérletezés közben (ez természetesen nem ajánlható az online közvetítőknek...) megbeszéli a felhasználóval a témát, a keresési szinonímákat. A képernyőn ellenőrzik az eredményt, a szükségeseinél nagyobb "szemét" esetén negatív kizárással is élnek (magyarán: ami nem kell az olvasónak, azt leválasztják...). A végső pontosítást az említett *free text searching* lehetősége biztosítja. Végül lementjük az immár végleges keresési találatokat. Esetleg több kinyomtatási formátum (PFT) is szóba kerülhet, maga a rendezés a sort print menüben már menüeditorral eleve kitöltött (rendezési kulcsok, nyomtató vezérlés stb.) paraméterekkel történik. Nevezetesen a rekordok — kiemelt fejléccel — újságok, folyóiratok szerint, ezen belül évjárat, majd végül szerzői rendben kerülnek kinyomtatásra. Ezt a rendezést a tapasztalat indokolja, hiszen az olvasók többsége a kinyomtatott irodalomjegyzékkel általában a folyóiratszolgálathoz megy, ahol így kényelmes és természetes sorrendben kereshet akár a bekötött anyagban, és kérhet végül fénymásolatokat is.

Nagyobb keresésekre a tanév során havonta 15-20 alkalommal kerül sor. A találati rekordok száma akár elérheti a 4-600-at is (ez 35-40 kinyomtatott oldal), de az átlagos keresés 100 rekord alatt marad. A szolgáltatás hallgatóknak és oktatóknak természetesen ingyenes, a külső érdeklődés pedig viszonylag kicsi.

Fejlődési lehetőségek

Apró, de figyelemre méltó változás a PRESSDOK "szegedi életében", hogy a tanszék vagy oktató nem nyomtatott anyagot kért, hanem letöltést floppy lemezre, sőt volt példa már arra is, hogy a rövidebb keresési eredményt az IIF levelező rendszerén (ELLA) keresztül juttattunk el a felhasználó gépébe.

A szolgáltatás haszna az olvasók, hallgatók és kutatók számára nyilvánvaló. Több tucatra tehető azon — remélhetőleg alapos — referátumok, diplomamunkák és disszertációk száma, amelyeknek előzetes irodalomkutatásában fontos szerepet játszott az ISIS alatt kumulált PRESSDOK. De a haszon a másik oldalon, a könyvtárban is jelentkezett. Többek között éppen ezen szolgáltatás itt vázolt útja teremtette meg a tapasztalatokat a hetvenes évek eleje óta készülő online nagykatalógusunk Micro-ISIS alapú létrehozása számára.

Jegyzetek

1. A nagykönyvtárban folyó számítógépesítési munkálatokról KOKAS Károly: Számítógépes adatfeldolgozás a Szegedi Egyetemi Könyvtárban. = Csongrád Megyei Könyvtáros. 22. évf. 1990. 1-2. sz. 32-43.p.
2. RÓNAI Iván: Számítógépesítés az Országgyűlési Könyvtárban. = Könyvtári Figyelő. 35. évf. 1989. 5-6. sz. 483-490.p.